

Optimization for Artificial Intelligence

Robustness, Overfitting, Transfer, Frugality.

Where ?

Ens Lyon 46 allée d'Italie, 69007 Lyon

- Amphi B de 8h30 à 12h30.
- Salle des thèses de 13h45 à 17h30

When ? February 14th, 2025.

Program

Who ?	What ?	When ?
Amphi B, 46 allée d'Italie, 69007 Lyon		
	Introduction	9:00 - 9:10
O. Teytaud	Optimization for Post-Training in Large Language Models	9:10 - 10:00
J. Tacchella	UNSURE: Unknown Noise level Stein's Unbiased Risk Estimator	10:00 - 10:50
	Coffee Break	
Cédric Gerbelot	Mean field theory for SGD “High-dimensional optimization for the multi-spike tensor PCA problem”	11:10-12:00
	Lunch break	
Salle des Theses, 46 allée d'Italie, 69007 Lyon		
R. Gribonval	Conservation laws during neural network training	14:00-14:50
E. Ricietti	Frequency-aware training of PINNs : a multigrid idea	14:50-15:40
G. Charpiat	Neural Network Growth	15:40-16:30
	The end !	16:30

Abstract de J. Tacchella: Recently, many self-supervised learning methods for image reconstruction have been proposed that can learn from noisy data alone, bypassing the need for ground-truth references. Most existing methods cluster around two classes: i) Stein's Unbiased Risk Estimate (SURE) and similar approaches that assume full knowledge of the distribution, and

ii) Noise2Self and similar cross-validation methods that require very mild knowledge about the noise distribution. The first class of methods tends to be impractical, as the noise level is often unknown in real-world applications, and the second class is often suboptimal compared to supervised learning. In this talk, I will present a theoretical framework that characterizes this expressivity-robustness trade-off and propose a new approach based on SURE, but unlike the standard SURE, does not require knowledge about the noise level. I will also show that the proposed estimator outperforms other existing self-supervised methods on various imaging inverse problems.

Abstract de C. Gerbelot: A core difficulty in modern machine learning is to understand the convergence of gradient based methods in random, high-dimensional, non-convex landscapes. In this work, we study the behavior of gradient flow and online stochastic gradient descent applied to the multi-spike tensor PCA problem, the goal of which is to recover a set of spikes from noisy observations of the corresponding tensor. The main thrust of our proof relies on a sharp control of the random part of the dynamics, followed by the analysis of a finite dimensional dynamical system, leading to both sample complexity bounds and a complete description of the set of critical points reached by the dynamics. In particular, we obtain sufficient conditions for reaching the global minimizer of the problem from uninformative initializations. At a technical level, we will put our methods, originating in probability and mathematical physics, in perspective with those used in machine learning theory and statistical physics of learning. This talk is based on joint work with Vanessa Piccolo and Gérard Ben Arous.

Abstract de E. Riccietti: It is well known that the training of physics-informed neural networks (PINNs) may be difficult and slow in the presence of high-frequencies. Interestingly, the same phenomenon arises in the solution of partial differential equations by classical smoothing methods, just in the opposite direction: the low frequencies are difficult to reduce. Multigrid methods address this challenge by exploiting the complementarity between the involved sub-problems and achieve great acceleration and computational savings. In this work we propose an extension of the basic principle of multigrid methods to the training of PINNs to obtain a frequency-aware training scheme. We show that our approach is particularly effective if coupled with specialized frequency-aware network architectures.